

This is a repository copy of *Evolving Graphs with Semantic Neutral Drift*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/154735/>

Version: Published Version

---

**Article:**

Plump, Detlef [orcid.org/0000-0002-1148-822X](https://orcid.org/0000-0002-1148-822X), Atkinson, Timothy and Stepney, Susan [orcid.org/0000-0003-3146-5401](https://orcid.org/0000-0003-3146-5401) (2021) *Evolving Graphs with Semantic Neutral Drift*. *Natural Computing*. pp. 127-143. ISSN 1567-7818

<https://doi.org/10.1007/s11047-019-09772-4>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



# Evolving graphs with semantic neutral drift

Timothy Atkinson<sup>1</sup> · Detlef Plump<sup>1</sup> · Susan Stepney<sup>1</sup>

© The Author(s) 2019

## Abstract

We introduce the concept of Semantic Neutral Drift (SND) for genetic programming (GP), where we exploit equivalence laws to design semantics preserving mutations guaranteed to preserve individuals' fitness scores. A number of digital circuit benchmark problems have been implemented with rule-based graph programs and empirically evaluated, demonstrating quantitative improvements in evolutionary performance. Analysis reveals that the benefits of the designed SND reside in more complex processes than simple growth of individuals, and that there are circumstances where it is beneficial to choose otherwise detrimental parameters for a GP system if that facilitates the inclusion of SND.

**Keywords** Genetic programming · Evolutionary algorithms · Neutral drift · Semantic equivalence · Mutation operators · Graph programming

## 1 Introduction

In genetic programming the ability to escape local optima is key to finding globally optimal solutions. Neutral drift, a mechanism whereby individuals with fitness-equivalent phenotypes to the existing population may be generated by mutation (Galván-López et al. 2011) offers the search of new neighborhoods for sampling thus increasing the chance of leaving local optima. A number of studies on neutrality in Cartesian Genetic Programming (CGP) (Miller and Smith 2006; Vassilev and Miller 2000; Turner and Miller 2015b) find it to be an almost always beneficial property for studied problems. In general, comparative studies (Miller 2011) find that CGP using only mutation and neutral drift is able to compete with traditional tree-based Genetic Programming (GP) which uses more familiar crossover operators (see Koza 1992) to introduce genetic variation.

Turner and Miller (2015b) makes a distinction between *implicit* neutral drift (where a genetic operator yields a semantically equivalent child) and *explicit* neutral drift (where a genetic operator only modifies intronic code). We note that many comparative studies largely focus on the role of both types of neutral drift as byproducts of existing genetic operators and neutrality within the representation (Miller and Smith 2006; Vassilev and Miller 2000; Turner and Miller 2015b; Banzhaf 1994) rather than as deliberately designed features of an evolutionary system. We propose the opposite; to employ domain knowledge of equivalence laws to specify mutation operators on the active components of individuals which always induce neutral drift. Hence our work can be viewed as an attempt to explicitly induce additional implicit neutral drift in the sense of Turner and Miller (2015b).

We build on our approach EGGP (Evolving Graphs by Graph Programming) (Atkinson et al. 2018a), by implementing *semantics preserving mutations* to directly achieve neutral drift on the active components of individual solutions. Here, we implement logical equivalence laws as mutations on the active components of candidate solutions to digital circuit problems to produce semantically equivalent, equally fit, children. While our semantics-preserving mutations produce semantically equivalent children they do not guarantee preservation of size; our fitness measures evaluate semantics only, not, for example, size or complexity.

---

✉ Timothy Atkinson  
tja511@york.ac.uk

Detlef Plump  
detlef.plump@york.ac.uk

Susan Stepney  
susan.stepney@york.ac.uk

<sup>1</sup> Department of Computer Science, University of York, York, UK

We describe and implement Semantic Neutral Drift straightforwardly by using rule-based graph programs, here in the probabilistic language P-GP 2 (Atkinson et al. 2018b). This continues from Atkinson et al. (2018a) where we use a probabilistic variant of the graph programming language GP 2 to design acyclicity-preserving edge mutations for digital circuits that correctly identify the set of all possible valid mutations. The use of P-GP 2 here enables concise description of complex transformations such as DeMorgan's laws by identifying and rewriting potential matches for these laws in the existing formalism of graph transformation. This reinforces the notion that the direct encoding of solutions as graphs is useful as it allows immediate access to the phenotype of individual solutions and makes it possible to design complex mutations by using powerful algorithmic concepts from graph programming.

We investigate four sets of semantics-preserving mutations for digital circuit design, three built upon logical equivalence laws and a fourth taken from term-graph rewriting. We run EGGP with each rule-set on a set of benchmark problems and establish statistically significant improvements in performance for most of our visited problems. An analysis of our results reveals evidence that it is the semantic transformations, beyond simple 'neutral growth', which are aiding performance. We then combine our two best performing sets of mutation operators and evaluate this new set under the same conditions, achieving further improvements. We also provide evidence that, although operators implementing semantics-preserving mutations may be more difficult to use, the inclusion of those semantics-preserving mutations may allow evolution to out-perform equivalent processes that use 'easier' operators.

The rest of this paper is organised as follows. In Sect. 2 we review existing literature on Neutral Drift in Genetic Programming. In Sects. 3 and 4 we describe the graph programming language GP 2 and our existing approach EGGP. In Sect. 5 we describe our extension to EGGP where we incorporate deliberate neutral drifts into the evolutionary process. In Sect. 6 we describe our experimental setup and in Sect. 7 we give the results from these experiments. In Sect. 8 we provide in-depth analysis of these results to establish precisely what components of our approach are aiding performance. In Sect. 9 we conclude our work and propose potential future work on this topic.

## 2 Neutral drift in genetic programming

Neutral drift remains a controversial subject in Evolutionary Computation. See Galván-López et al. (2011) for a survey. Here, we focus on neutrality in the context of

genetic programming as the most relevant area to our own work; there is also literature on, for example, genetic algorithms (Harvey and Thompson 1997) and landscape analysis (Barnett 1998).

The process of neutral drift might be described as the mutation of individual candidate solutions to a given problem without advantageous or deleterious effect on their fitness. This exposes the evolutionary algorithm to a fitness 'plateau' with each fitness-equivalent individual offering a different portion of the landscape to sample. Neutral drift can be viewed as random walks on the neighborhoods of surviving candidate solutions. In a system with neutral drift, an apparently local optimum might be escaped by 'drifting' to some other fitness-equivalent solution that has advantageous mutations available.

The most apparent demonstration of neutral drift in genetic programming literature occurs in Cartesian Genetic Programming (CGP) (Miller and Thomson 2000). In CGP, individuals encode directed acyclic graphs; some portion of a genome may be 'inactive', contributing nothing to the phenotypic fitness, because it represents a subgraph that is not connected to the phenotype's main graph. These inactive genes can mutate without influencing an individual's fitness and then, at some later point, may become active. Early work on CGP has found that by allowing neutral drift to take place (by choosing a fitness-equivalent child over its parent in the  $1 + \lambda$  algorithm), the success rate of experiments significantly improves (Vassilev and Miller 2000). A later claim that neutrality in CGP aids search in needle-in-haystack problems (Yu and Miller 2002) has been contested by a counter-claim that better performance can be achieved by random search (Collins 2006). Miller and Smith (2006) finds that better performance can be achieved with neutral drift enabled by increasing the amount of redundant material present in individuals. Turner and Miller (2015b) establishes a distinction between *explicit* and *implicit* neutral drift. Explicit neutral drift occurs on inactive components of the individual, whereas implicit neutral drift occurs when active components of the individual are mutated but the fitness does not change. The authors were able to isolate explicit neutral drift and demonstrate that it offers additive benefits beyond those of implicit neutral drift.

Outside of CGP, Banzhaf (1994) proposes a form of Linear Genetic Programming where programs are decoded from bit-strings, and redundancy exists, in that certain operations have multiple representations. A study of evolvability in Linear GP (Hu and Banzhaf 2009) found that neutrality cooperates with 'variability' (the ability of a system to generate phenotypic changes) to generate adaptive phenotypic changes which aid the overall ability of the system to respond to the landscape. Recent work Hu and Banzhaf (2018) studying the role of neutrality in small

Linear GP programs found that the robustness of a genotype (the proportion of its neighbours within the landscape which are neutral changes) has a complex and non-monotonic relationship with the overall evolvability of the genotype.

In Downing (2005), binary decision diagrams are evolved with explicit neutral mutations. Although those neutral mutations are not isolated for their advantages/disadvantages, a later work has found that a higher rate of neutral drift on binary decision diagrams is advantageous (Downing 2006). Koza also makes some reference to the ideas we employ in Sect. 5 when he describes the editing digital circuits by applying DeMorgan's laws to them (Koza 1992, Ch. 6). A study of neutrality in tree-based GP for boolean functions (Vanneschi et al. 2012) found a correlation between using a more effective function set and the existence of additional neutrality when using that function set.

While not directly related to neutrality, a number of investigations have been carried out exploring the notion of semantically aware genetic operators to improve the locality of mechanisms such as crossover in tree-based GP (Moraglio et al. 2012; Nguyen et al. 2009). We refer the reader to the extensive survey (Vanneschi et al. 2014) on this field of research. Whereas neutrality is the process whereby phenotypically identical and genotypically distinct individuals are visited by the evolutionary process, semantically aware genetic operators attempt to produce phenotypically 'close' individuals to improve the locality of the search neighbourhood. It should be noted that employing semantically aware genetic operators may sometimes lead to a loss of diversity (Pham et al. 2013). It could be argued that the deliberate neutral operators we propose in this work are a form of semantically aware mutation operators designed to explicitly exploit neutrality.

Neutral drift has some parallels with work on biological evolution. Kimura's *Neutral Theory of Molecular Evolution* (Kimura 1983) posits that most mutations in nature are neither advantageous or deleterious, instead introducing 'neutral' changes that do not affect phenotypes but account for much of the genetic variation within and between species. While Kimura's theory remains controversial (see Hahn 2007), it appears to loosely correspond to the notions of neutral mutation described in genetic programming literature.

Throughout the literature we have covered, neutrality is mostly considered in the sense of *explicit neutral drift* as defined in Turner and Miller (2015b). Conversely in our work here we are focusing on neutral drift on the active components of individual solutions, with some relationship therefore to the neutral mutations on binary decision diagrams in Downing (2005).

### 3 Graph programming with P-GP 2

Here we give a brief introduction to the graph programming language GP 2; see Plump (2017) for a detailed account of the syntax and semantics of the language.

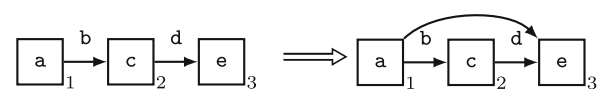
A graph program consists of declarations of *graph transformation rules* and a main command sequence controlling the application of the rules. Graphs are directed and may contain loops and parallel edges. The rules operate on *host graphs* whose nodes and edges are labelled with integers, character strings or lists of integers and strings. Variables in rules (relevant for this paper) are of type *int*, *string* or *list*. Integers and strings are considered as lists of length one, hence every label in GP 2 is a list. For example, in Fig. 1, the list variables *a*, *c* and *e* are used as node labels while *b* and *d* serve as edge labels. The small numbers attached to nodes are identifiers that specify the correspondence between the nodes in the left and the right graph of the rule.

Besides carrying list expressions, nodes and edges can be *marked*. For example, in the program of Fig. 3, blue and red node marks are used to prevent the rule *mutate\_edge* from creating a cycle. In rules, a magenta colour can be used as a wildcard for any mark. For example, in the rules *remove\_edge*, *unmark\_edge* and *unmark\_node* of Fig. 6, pairs of magenta nodes with the same identifier on the left and the right represent nodes with the same green, blue or grey mark.

The principal programming constructs in GP 2 are conditional graph-transformation rules labelled with expressions. To apply a rule to a host graph, the rule is first instantiated by replacing all variables with values and evaluating the expressions. The rule's condition, if present, has to evaluate to true. Then the left graph of the instantiated rule is matched (injectively) with a subgraph of the host graph. Finally the subgraph is replaced with the right graph of the instantiated rule. This means that the nodes corresponding to the numbered nodes of the left graph are preserved (but possibly re-labelled), any other nodes and all edges of the left graph are deleted, and any unnumbered nodes and all edges of the rule's instantiated right graph are inserted.

```
Main := link!
```

```
link(a,b,c,d,e:list)
```



```
where not edge(1 3)
```

Fig. 1 A GP 2 program computing the transitive closure of a graph

For example, given any host graph  $G$ , the program in Fig. 1 produces the smallest transitive graph that results from adding unlabelled edges to  $G$ . (A graph is *transitive* if for each directed path from a node  $v_1$  to another node  $v_2$ , there is an edge from  $v_1$  to  $v_2$ .) The program applies the single rule *link as long as possible* to a host graph. In general, any subprogram can be iterated with the postfix operator “!”. Applying *link* amounts to non-deterministically selecting a subgraph of the host graph that matches *link*’s left graph, and adding to it an edge from node 1 to node 3 provided there is no such edge (with any label). The application condition `where not edge(1,3)` ensures that the program terminates and extends the host graph with a minimal number of edges.

Besides applying individual rules, a program may apply a rule set  $\{r_1, \dots, r_n\}$  to the host graph by non-deterministically selecting a rule  $r_i$  among the applicable rules and applying it. Further control constructs include the sequential composition  $P; Q$  of programs  $P$  and  $Q$ , and the branching construct `try T then P else Q`. To execute the latter, test  $T$  is executed on the host graph  $G$  and if this results in some graph  $H$ , program  $P$  is executed on  $H$ . If  $T$  fails (because a rule or set of rules cannot be matched), program  $Q$  is executed on  $G$ . The variant `try T` of this construct executes  $T$  on  $G$  and if this results in graph  $H$ , returns  $H$ . If the execution fails,  $G$  is returned unmodified.

In general, the execution of a program on a host graph may result in different graphs, fail, or diverge. The *semantics* of a program  $P$  maps each host graph to the set of all possible outcomes (Plump 2012). GP 2 is computationally complete in that every computable function on graphs can be programmed (Plump 2017).

GP 2’s inherent non-determinism is useful as many graph problems are naturally multi-valued, for example the computation of a shortest path or a minimum spanning tree. The results described in the rest of this paper have been obtained with a probabilistic extension of GP 2, called P-GP 2. This provides a rule-set command  $[r_1, \dots, r_n]$  which chooses a rule uniformly at random among the applicable rules and applies the rule with a match selected uniformly at random among all matches of that rule (Atkinson et al. 2018b).

## 4 Evolving graphs by graph programming (EGGP)

### 4.1 Introduction to EGGP

In Atkinson et al. (2018a) we introduce EGGP, an evolutionary algorithm that evolves graphs (specifically, in that case, digital circuits) using graph programming. We have found that by evolving graphs directly and designing

mutation operators that respect the constraints of the problem, we are able to significantly outperform CGP under similar conditions on a number of digital circuit benchmark problems. In this section we formally describe this approach.

Our approach is justified by two observations: (1) the use of graphs as a representation is beneficial, as it directly addresses a number of motivating problems within computer science such as neural network topology, Bayesian network topology, digital circuit design, program design, and quantum circuit design; (2) with graphs as a representation it is necessary to have a language to describe the neighborhoods (mutations) on individuals. Graph programming readily lends itself to this endeavour due to its computational completeness over functions on graphs.

Our approach is not alone in addressing the issue of evolving graphs and graph-like programs. CGP (Miller and Thomson 2000), where individuals encode directed acyclic graphs, is a primary candidate for related work and is used as a benchmark here. Parallel Distributed Genetic Programming (Poli 1997, 1999) introduces a ‘graph on a grid’ representation for genetic programming in a similar manner to the grid-like description of CGP, allowing the evolution of programs with multiple outputs and sharing. MIOST (López et al. 2007) also extends traditional genetic programming to these same concepts of multiple outputs and sharing. For a more detailed discussion of related approaches, see Atkinson et al. (2018a). Our approach differs from these in that (1) we deal with graphs directly rather than through an encoding or some subset of graphs; and (2) our mutation operators are domain-specific and may be changed to suit the constraints of a problem and to exploit domain-specific knowledge.

Here we address the problems of digital circuits, primarily because they suit our discussion of neutral drift by design. For this reason, the rest of this paper focuses on the evolution of digital circuits as a concrete case study.

### 4.2 Evolving digital circuits as graphs

We directly encode digital circuits as graphs such that the graph contains input and output nodes (corresponding to the inputs and outputs of the intended problem) and function nodes. In P-GP 2, we identify input nodes and output nodes by labels of the form  $\text{IN} : x$  and  $\text{OUT} : y$  respectively, where  $x$  and  $y$  are integers that identify which particular input or output the node corresponds to. Function nodes are labelled as  $[f_i] : a$ , where  $[f_i]$  is a string uniquely identifying function  $f_i \in F$  and  $a$  is the arity of  $f_i$ . In this work our functions are symmetrical, but an extension is available to associate each edge with an integer to identify which particular input of a function it corresponds to. Figure 2 shows a digital circuit encoded in this form.

For a specific  $i$  input,  $o$  output problem over function set  $F$ , we must evolve graphs that are constrained:

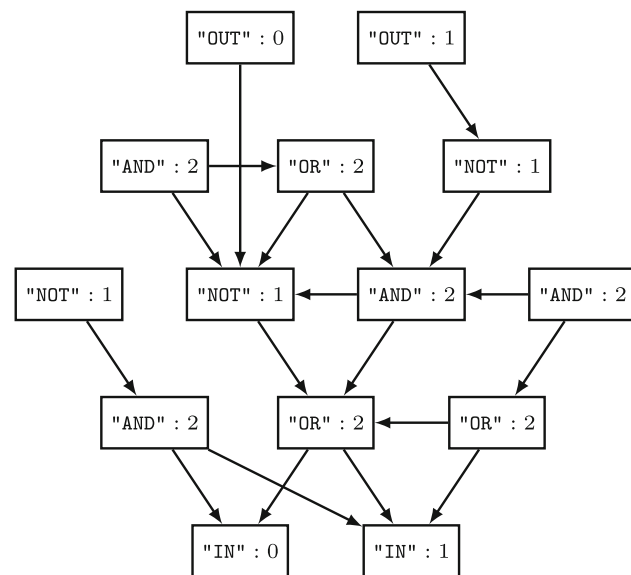
- Individual solutions are acyclic.
- Individual solutions have  $i$  input nodes.
- Individual solutions have  $o$  output nodes.
- All other nodes that are neither inputs nor outputs must be function nodes associated with some function  $f_i \in F$  and have exactly  $a$  outgoing edges where  $a$  is the arity of  $f_i$ .

We use three graph programs to induce a landscape; `InitCircuit`, `MutateFunction` and `MutateEdge`. The first is the initialisation program for generating individual graphs, and the others are mutation operators. `InitCircuit` and `MutateFunction` are given in “Appendix”; it should be clear that they satisfy the constraints described above. Here we describe in more detail the `MutateEdge` operator, which is the mutation operator primarily responsible for the topological changes to individual solutions.

The `MutateEdge` operator is shown in Fig. 3. It works by first picking an edge to mutate at random using the `pick_edge` rule, marking that edge red, its source blue and its target red. Then `mark_output` is applied as long as possible, marking blue every node for which there is a directed path to the source of the edge we wish to mutate. `mutate_edge` can be safely applied to redirect the edge to

target some unmarked node (chosen at random); this cannot introduce a cycle as the new target is unmarked and therefore does not have a directed path to the existing source of the mutating edge. Finally `unmark` is applied as long as possible to return the graph to an unmarked state. This P-GP 2 program uses a uniform random distribution to chose the edge to mutate, a uniform distribution over all possible edge mutations that preserve acyclicity, and clearly respects the other constraints mentioned above, as it does not relabel any nodes or change the number of outgoing edges of any node. In Atkinson et al. (2018a) we argue that this edge mutation generalises the order preserving mutations of CGP and offers additional possible mutations. A visual step-by-step execution of this mutation operator is shown in Fig. 4.

In general, we use the  $1 + \lambda$  evolutionary algorithm with EGGP.  $1 + \lambda$  has been used extensively with CGP with favourable comparisons with large-population GP systems (see Miller 2011). A comparative study of crossover in CGP (Husa and Kalkreuth 2019) found that there is no currently known universal crossover operator for CGP and that  $1 + \lambda$  is sometimes the best known approach for certain problems. Current advice (Miller 2011; Turner and Miller 2015a) is to use  $1 + \lambda$  as the ‘standard’ CGP approach. The comparative study between EGGP and CGP (Atkinson et al. 2018a) exclusively used the  $1 + \lambda$  strategy with EGGP performing favourably on many digital circuit benchmark problems. In combination, these points appear to justify the exclusive use of  $1 + \lambda$  with EGGP in our study. Additionally, the use of  $1 + \lambda$  has the added effect of ‘isolating’ our notion of semantic neutral drift, in that we can apply logical equivalence laws to the single surviving individual in each generation knowing that its application is not disrupting other processes e.g. crossover or non-elitist selection.

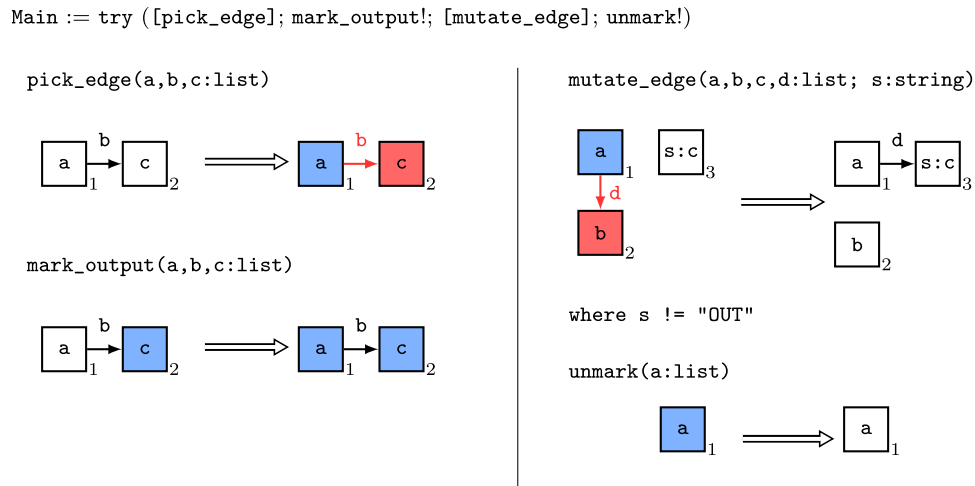


**Fig. 2** A P-GP 2 encoding of a 2-input, 2-output digital circuit over the function set {AND, OR, NAND, NOR}. (The outgoing edges of a node point to the sources of that node’s input values, following the convention used in the graph programming community). Output 0 (corresponding to the node labelled }OUT} : 0) has logical behaviour  $\neg(i_0 \vee i_1)$  where  $i_0$  and  $i_1$  correspond to the input nodes labelled }IN} : 0 and }IN} : 1 respectively.

## 5 Semantic neutral drift

### 5.1 The concept

Semantic Neutral Drift (SND) is the augmentation of a GP system with semantics-preserving mutations. These mutations are added to the standard mutation and cross-over operators, which are intended to introduce variation to search. In this section we refer to mutation operators and individuals generally, not just our specific operation. For individual solutions  $i, j$  and mutation operator  $m$ , we write  $i \rightarrow_m j$  to mean that  $j$  can be generated from  $i$  by using mutation  $m$ . A semantics-preserving mutation is one that guarantees that the semantic meaning of a child generated by that mutation is identical to that of its parent, for any choice of parents and a given semantic model. This



**Fig. 3** A P-GP 2 edge mutation `MutateEdge` for digital circuits. This edge mutation preserves acyclicity. The rule `pick_edge` is used to probabilistically choose an edge to mutate. Then `mark_output` is applied as long as possible, marking every node with a path to the source of the edge we wish to mutate blue. `mutate_edge` can then be

applied safely, redirecting the edge to target some unmarked node which does not have a path to the source of the mutating edge. Finally `unmark` is applied as long as possible to return the graph to an unmarked state

definition is adequate for our domain of GP, where there is no distinction between the genotype and phenotype.

For our digital circuits case study, this semantic equivalence is well-defined: two circuits are semantically equivalent if they describe identical truth tables. Therefore, semantics preserving mutations in this context are ones which preserve an individual's truth table. As we will be evaluating individuals by the number of incorrect bits in their truth tables, there may be individuals with equivalent fitness but different truth tables. Therefore, semantic equivalence is distinct from, but related to, fitness equivalence.

Additionally, semantics preserving mutations do not necessarily induce neutral drift. In the circumstance that a fitness function considers more than the semantics of an individual, there is no guarantee that the child of a parent generated by a semantics-preserving mutation has equal fitness to its parent. For example, if a fitness function penalized the size of an individual, a semantics-preserving mutation which introduces additional material (e.g. increases size) would generate children less fit than their parents under this measure.

We identify a special class of fitness functions, where fitness depends only on semantics, and so where semantics-preserving mutations are guaranteed to preserve fitness. In this circumstance, any use of semantics-preserving mutations is a deliberate, designed-in, form of neutral drift. The fitness function in our case study is an example of this; the fitness of an individual depends only on its truth table. Formally we have the following: a set of semantics-preserving mutation operators  $M$  over search space  $S$  with

respect to a fitness function  $f$  that considers only semantics guarantees that

$$\forall i, j \in S, m \in M : (j \rightarrow_m i) \Rightarrow (f(i) = f(j)).$$

Consider a GP run that has reached a local optimum; no available mutations or crossover operators offer positive improvements with respect to the fitness function. It may be the case that there is a solution elsewhere in the landscape that is equally fit as the best found solution but has a neighborhood with positive mutations available. By applying a semantics preserving mutation to transform the best found solution into this other, semantically equivalent, solution, the evolutionary process gains access to this better neighborhood and can continue its search. Hence the proposed benefit of Semantic Neutral Drift is the same as conventional neutral drift: that by transforming discovered solutions we gain access to different parts of the landscape that may allow the population to escape local optima. The distinction here is that we are employing domain knowledge to deliberately preserve semantics, rather than accessing neutral drift as a byproduct of other evolutionary processes. The hypothesis we are investigating is that this deployment of domain knowledge yields more meaningful neutral mutations than simple rewrites of intronic code, and that this leads the evolutionary algorithm to more varied (and therefore useful) neighborhoods.

A simple visualization of Semantic Neutral Drift is given in Fig. 5. Here the landscape exists in one dimension (the  $x$ -axis) with fitness of individuals given in the  $y$ -axis. In this illustration, the individual has reached a local optimum, then a semantics-preserving mutation moves it to a

This individual is to undergo an edge mutation preserving acyclicity.

(1) `pick_edge`:  
An edge to mutate is chosen at random and marked (red) alongside its source node  $s$  (blue) and target node  $t$  (red).

(2) `mark_output!`:  
Invalid candidate nodes for redirection are identified. If a node  $v$  has a directed path to  $s$  it is marked blue, as targeting it would introduce a cycle.

(3) `mutate_edge; unmark!`:  
The edge  $e$  is mutated to target some randomly chosen unmarked (non-output) node, preserving acyclicity. The new target has been marked with an star '\*' for visual clarity. Finally, all marks are removed.

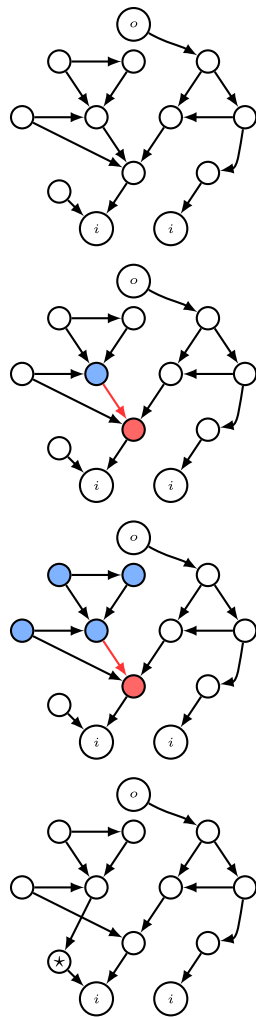


Fig. 4 A step-by-step execution of the edge mutation operator given in Fig. 3. For visual simplicity, node labels have been omitted

different ‘hill’ from which it is able to reach the global optimum.

While our experiments will focus on the role of semantic neutral drift when evolving graphs with EGGP, we argue that the underlying concept is extendable to other GP systems. For example, Koza noted the possibility of applying DeMorgan’s laws to GP trees (Koza 1992, Ch. 6) which, if used in a continuous process rather than as a solution optimiser, would induce semantic neutral drift. It is also plausible to apply similar operators to CGP (Miller and Thomson 2000) representations, although the ordering imposed on the representation raises some technical difficulties with respect to where newly created nodes should be placed. The potential for Embedded CGP (Walker and Miller 2008) to effectively grow and shrink the overall size of the genotype offers some hope in this direction.

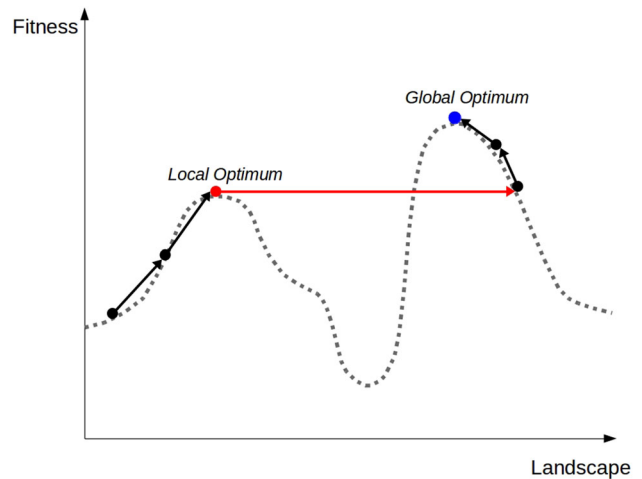


Fig. 5 A simple visualization of Semantic Neutral Drift. Individuals exist in one dimension along the  $x$ -axis. On the  $y$ -axis, each individual has an associated fitness. Normal mutations (black arrows) allow the evolutionary algorithm to hill-climb by sampling from adjacent points. A semantics-preserving mutation (red arrow) allows the EA to leave a local optimum to move to a different slope where it can then climb to the global optimum. (Color figure online)

### 5.2 Designing semantic neutral drift

We extend EGGP by applying semantics-preserving mutations to members of the population each generation. Here we focus on digital circuits as a case study, and design mutations which modify the *active* components of the individual by exploiting domain knowledge of logical equivalence.

For the function set {AND, OR, NOT} there are a number of known logical equivalences. Here we use DeMorgan’s laws:

$$\begin{aligned} \text{DeMorgan}_{F1} &: \neg(a \wedge b) = \neg a \vee \neg b \\ \text{DeMorgan}_{F2} &: \neg(a \vee b) = \neg a \wedge \neg b \\ \text{DeMorgan}_{R1} &: \neg a \vee \neg b = \neg(a \wedge b) \\ \text{DeMorgan}_{R2} &: \neg a \wedge \neg b = \neg(a \vee b) \end{aligned}$$

and the identity and double negation laws:

$$\begin{aligned} \text{ID-AND}_F &: a = a \wedge a \\ \text{ID-AND}_R &: a \wedge a = a \\ \text{ID-OR}_F &: a = a \vee a \\ \text{ID-OR}_R &: a \vee a = a \\ \text{ID-NOT}_F &: a = \neg\neg a \\ \text{ID-NOT}_R &: \neg\neg a = a \end{aligned}$$

Here we investigate different subsets of these semantics-preserving rules. We encode them as graph transformation rules to apply to the active component of an individual. In the context of the  $1 + \lambda$  evolutionary algorithm, we apply



one of the rules from the subset to the surviving individual of each generation.

Encoding these semantics-preserving rules is non-trivial for our individuals as they incorporate sharing; multiple nodes may use the same node as an input, and therefore rewriting or removing that node, e.g. as part of DeMorgan's, may disrupt the semantics elsewhere in the individual. To overcome this, we need a more sophisticated rewriting program. The graph program in Fig. 6 is designed for the logical equivalence laws  $DeMorgan_{F1|F2}$ ,  $DeMorgan_{R1|R2}$ ; analogous programs are used for other operators.

The program Main in Fig. 6 works as follows.

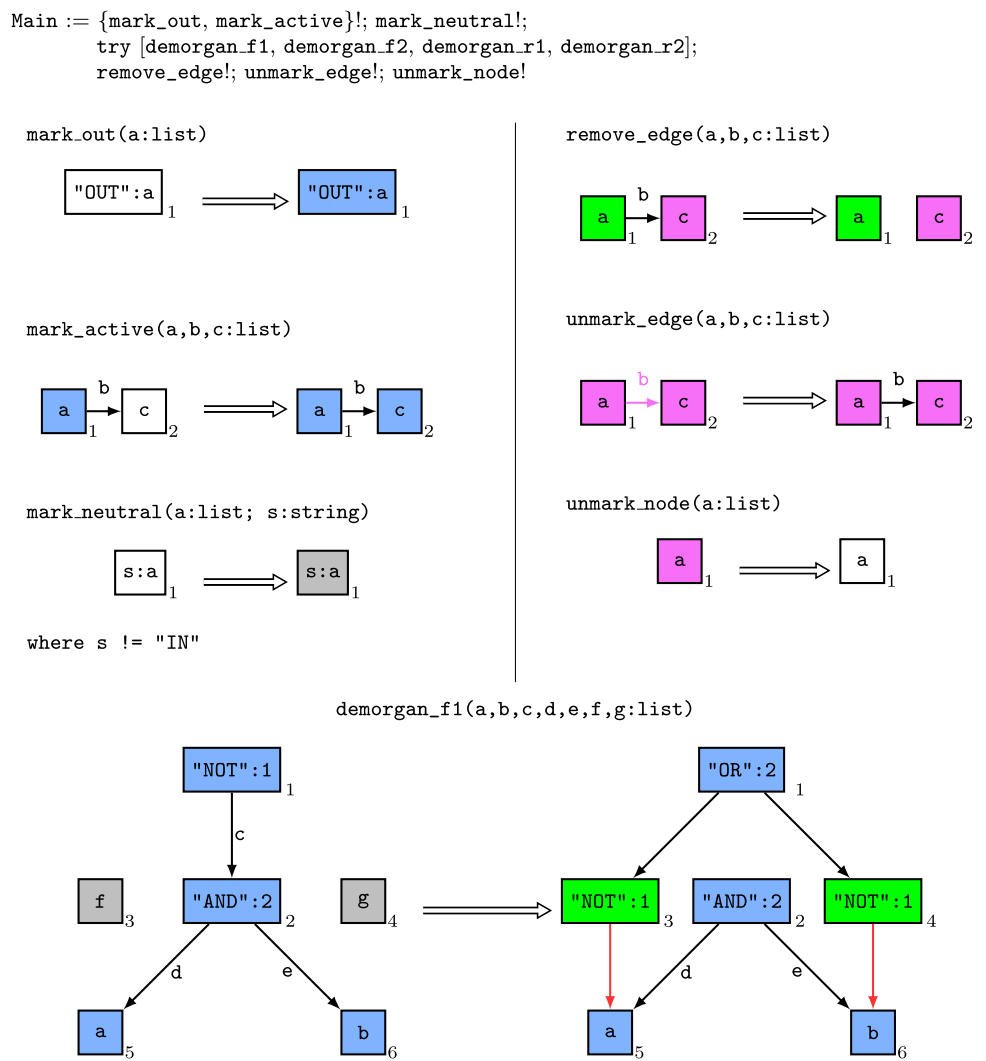
`{mark_out, mark_active}!`: Mark all active nodes with the given rule-set applied as long as possible. Once this rule-set has no matches, all inactive nodes must be unmarked: these are 'neutral' nodes that do not contribute to the semantics of the individual.

`mark_neutral!`: Mark these neutral nodes grey with the rule applied as long as possible. We can then rewrite the individual while preserving semantics with respect to shared nodes by incorporating neutral nodes into the active component rather than overwriting existing nodes.

`try [demorgan_f1, demorgan_f2, demorgan_r1, demorgan_r2]`: pick some rule with uniform probability from the subset of the listed rules that have valid matches. When a rule has been chosen, a match is chosen for it from the set of all possible matches with uniform probability. The probabilistic rule-set call is surrounded by a `try` statement to catch the fail case that none of the rules have matches.

In Fig. 6 we show one of the 4 referenced rules, `demorgan_f1`, which corresponds to the logical equivalence law  $DeMorgan_{F1}$ ; the others may be given analogously. On the left hand side is a match for the pattern  $\neg(a \wedge b)$  in the active component and 2 neutral nodes. If the matched pattern were directly transformed, any nodes

**Fig. 6** A P-GP 2 program for performing semantics preserving mutations to digital circuits



sharing use of the matches for node 2 or node 3 could have their semantics disrupted. Instead, the right-hand-side of demorgan\_f1 changes the syntax of node 1 to correspond to  $\neg a \vee \neg b$  by absorbing the matched neutral nodes (preserving its semantics) without rewriting nodes 1 or 2 and disrupting their semantics. Nodes 3 and 4 are marked green and their newly created outgoing edges are marked red. These marks are used later in the program to clean up any previously existing outgoing edges they have to other parts of the graph.

`remove_edge`: once a semantics preserving rule has been applied, the rule is applied as long as possible to remove the other outgoing edges of green marked absorbed nodes.

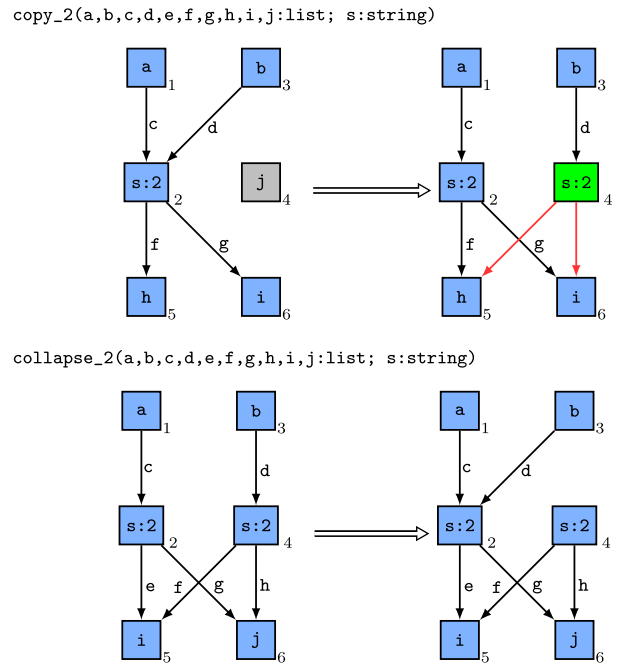
`unmark_edge!`; `unmark_node!`: return the graph to an unmarked state, where nodes and edges with any mark (indicated by magenta edges and nodes in the rules) have their marks removed.

This program highlights the helpfulness of graph programming for this task. The probabilistic application of complex transformations, such as DeMorgan’s law, to only the active components of a graph-like program with sharing is non-trivial, but can be concisely described by a graph program.

### 5.3 Variations on our approach

We identify 3 sets of logical equivalence rules to study, alongside another example of semantics preserving transformation taken from term-rewriting theory. These sets are detailed in Table 1. The first 3 sets comprise the logical equivalence laws already discussed. The last, CC, refers to collapsing and copying from term graph rewriting (see Habel et al. 1988). Collapsing is the process of merging semantically equivalent subgraphs, and copying is the process of duplicating a subgraph.

The rules `collapse2` and `copy2` are shown in Fig. 7. These collapse and copy, respectively, function nodes of arity 2 without garbage collection. We only require rules for arity 1 and arity 2 as our function sets in experiments are limited to arity 2. This final set is included for several reasons: it takes a different form from the domain-specific logical equivalence laws in the other 3 sets; it allows us to



**Fig. 7** The rules `copy2` and `collapse2`. The rule `copy2` matches a 2-arity function node that is shared by 2 active nodes and absorbs a neutral node to effectively copy that 2-arity function node and redirect one of the original node’s shared incoming edges to that copy. The rule `collapse2` attempts the reverse of `copy2` by matching 2 active identical 2-arity function nodes and redirecting one of those nodes’ incoming edges to the other. The node which has lost an incoming edge, if it was shared by no other nodes, may now become neutral.

investigate if the apparent overlap between term-graph rewriting and evolutionary algorithms bears fruit; it appears to resemble gene duplication, which is a natural biological process believed to aid evolution (Zhang 2003).

## 6 Experimental setup

To evaluate our approach, we study the same digital circuit benchmark problems as in Atkinson et al. (2018a), listed in Table 2. We perform 100 runs of each of our 4 neutral drift sets (Table 1) on each problem (Table 2). We use the  $1 + \lambda$  evolutionary algorithm with  $\lambda = 4$ . We use a mutation rate

**Table 1** The studied semantics preserving rule-sets

| Set                         | Rules   |
|-----------------------------|---|
| DeMorgan (DM)               | DeMorgan <sub>F1</sub> , DeMorgan <sub>F2</sub> , DeMorgan <sub>R1</sub> , DeMorgan <sub>R2</sub>   |
| DeMorgan and Negation (DMN) | DeMorgan <sub>F1</sub> , DeMorgan <sub>F2</sub> , DeMorgan <sub>R1</sub> , DeMorgan <sub>R2</sub> , ID-NOT <sub>F</sub> , ID-NOT <sub>R</sub> |
| Identity (ID)               | ID-AND <sub>F</sub> , ID-AND <sub>R</sub> , ID-OR <sub>F</sub> , ID-OR <sub>R</sub> , ID-NOT <sub>F</sub> , ID-NOT <sub>R</sub>               |
| Collapse/Copy (CC)          | collapse <sub>1</sub> , collapse <sub>2</sub> , copy <sub>1</sub> , copy <sub>2</sub>   |

**Table 2** Digital circuit benchmark problems

| Digital circuit                | No. inputs | No. outputs |
|--------------------------------|------------|-------------|
| 1-bit adder (1-Add)            | 3          | 2           |
| 2-bit Adder (2-Add)            | 5          | 3           |
| 3-bit Adder (3-Add)            | 7          | 4           |
| 2-bit Multiplier (2-Mul)       | 4          | 4           |
| 3-bit Multiplier (3-Mul)       | 6          | 6           |
| 3:8-bit De-Multiplexer (DeMux) | 3          | 8           |
| 4 × 1-bit Comparator (Comp)    | 4          | 18          |
| 3-bit Even Parity (3-EP)       | 3          | 1           |
| 4-bit Even Parity (4-EP)       | 4          | 1           |
| 5-bit Even Parity (5-EP)       | 5          | 1           |
| 6-bit Even Parity (6-EP)       | 6          | 1           |
| 7-bit Even Parity (7-EP)       | 7          | 1           |

of 0.01 and fix all individuals to use 100 function nodes. The fitness function used is the number of incorrect bits in an individual's truth table compared to the target truth table, hence we are minimizing the fitness. We are able to achieve 100% success rate in finding global optima in our evolutionary runs, so we compare the number of evaluations required to find perfect fitness.

The function set used here is {AND, OR, NOT}, rather than the set {AND, OR, NAND, NOR} used in Atkinson et al. (2018a) and (Miller 2011, Ch.2). Our function set is chosen to directly correspond to the logical equivalence laws used. To give context to the results in Sect. 7, and to highlight that the chosen function set is the harder of the two, we run EGGP with both function sets and detail the results in Table 3. For additional context, the comparative study in Atkinson et al. (2018a) has shown EGGP to perform favourably in comparison to CGP on these problems with the {AND, OR, NAND, NOR} function set.

We use a two-tailed Mann–Whitney  $U$  test (Mann and Whitney 1947) to establish a statistically significant difference between the median number of evaluations using the two different function sets. When a result is statistically significant ( $p < 0.05$ ) we also use a Vargha–Delaney  $A$  test (Vargha and Delaney 2000) to measure the effect size. On every problem, using {AND, OR, NOT} takes significantly ( $p < 0.05$ ) more effort (in terms of evaluations) than when using {AND, OR, NAND, NOR}. This justifies our assertion that the former function set is 'harder' to evolve.

## 7 Results

The results from our experiments are given in Table 4. Each neutral rule-set is listed with the median evaluations (ME) required to solve each benchmark problem.

**Table 3** Baseline results from digital circuit benchmarks for EGGP on the {AND, OR, NOT} and {AND, OR, NAND, NOR} function sets. ME/IQR: the median/inter-quartile range of the number of evaluations used to solve the problem

| Problem | EGGP           |         |                      |         |            |             |
|---------|----------------|---------|----------------------|---------|------------|-------------|
|         | {AND, OR, NOT} |         | {AND, OR, NAND, NOR} |         | $p$        | $A$         |
|         | ME             | IQR     | ME                   | IQR     |            |             |
| 1-Add   | 15,538         | 18,963  | 7495                 | 8764    | $10^{-7}$  | 0.71        |
| 2-Add   | 162,003        | 172,781 | 82,688               | 79,333  | $10^{-8}$  | <b>0.73</b> |
| 3-Add   | 742,948        | 679,040 | 309,570              | 288,865 | $10^{-16}$ | <b>0.83</b> |
| 2-Mul   | 21,733         | 28,319  | 14,263               | 13,801  | $10^{-4}$  | 0.65        |
| 3-Mul   | 1,326,880      | 907,544 | 932,430              | 643,529 | $10^{-6}$  | 0.68        |
| DeMux   | 28,123         | 17,450  | 17,100               | 10,763  | $10^{-9}$  | <b>0.75</b> |
| Comp    | 408,448        | 275,581 | 147,343              | 128,304 | $10^{-17}$ | <b>0.85</b> |
| 3-EP    | 7403           | 8051    | 4295                 | 5500    | $10^{-4}$  | 0.66        |
| 4-EP    | 26,715         | 20,430  | 16,445               | 13,568  | $10^{-9}$  | <b>0.73</b> |
| 5-EP    | 76,608         | 57,518  | 42,778               | 29,454  | $10^{-10}$ | <b>0.75</b> |
| 6-EP    | 175,908        | 120,504 | 80,940               | 56,283  | $10^{-15}$ | <b>0.83</b> |
| 7-EP    | 380,600        | 237,965 | 157,755              | 118,065 | $10^{-19}$ | <b>0.87</b> |

The  $p$  value is from the two-tailed Mann–Whitney  $U$  test. Where  $p < 0.05$ , the effect size from the Vargha–Delaney  $A$  test is shown; large effect sizes ( $A > 0.71$ ) are shown in bold

We use a two-tailed Mann–Whitney  $U$  test to demonstrate statistical significance in the difference of the median evaluations for these runs and the unmodified EGGP results given in Table 3.

For most problems and neutral rule-sets, the inclusion of semantic neutral drift yields statistically significant improvements in performance. There are some exceptions: for the 4 × 1-bit comparator (COMP) problem, the inclusion of neutral rule-sets leads either to insignificant differences or to significantly worse performance for every rule-set except ID, which performs significantly better. The DeMorgan's rule-set (DM) and Copy/Collapse rule-set (CC) appear to yield the smallest benefit, finding significant improvement on only 8 and 9 of the 13 benchmark problems respectively. Additionally, both of these rule-sets yield significantly worse performance for the 4 × 1-bit comparator (COMP) problem. The DeMorgan's and Negation rule-set (DMN) offer the best performance on the 2-bit and 3-bit adder problems (2-Add and 3-Add), in terms of median evaluations,  $p$  value and effect size. The Identity rule-set (ID) achieves the best performance on the 2-bit and 3-bit multiplier problems (2-Mul and 3-Mul) but fails to achieve significant improvements on the 3:8-bit de-multiplexer problem (DeMux).

Our results show that, for some problems and certain neutral rule-sets, the inclusion of neutral drift may improve performance with respect to the effort (measured by the

**Table 4** Results from digital circuit benchmarks for the various proposed neutral rule-sets

| Circuit | Neutral ruleset |            |             |         |            |             |         |            |             |         |            |             |
|---------|-----------------|------------|-------------|---------|------------|-------------|---------|------------|-------------|---------|------------|-------------|
|         | DM              |            |             | DMN     |            |             | ID      |            |             | CC      |            |             |
|         | ME              | $p$        | $A$         | ME      | $p$        | $A$         | ME      | $p$        | $A$         | ME      | $p$        | $A$         |
| 1-Add   | 8950            | $10^{-7}$  | <b>0.72</b> | 9893    | $10^{-5}$  | 0.68        | 9093    | $10^{-7}$  | 0.71        | 8275    | $10^{-7}$  | <b>0.72</b> |
| 2-Add   | 65,692          | $10^{-14}$ | <b>0.81</b> | 49,200  | $10^{-21}$ | <b>0.88</b> | 73,275  | $10^{-12}$ | <b>0.79</b> | 103,393 | $10^{-5}$  | 0.68        |
| 3-Add   | 255,003         | $10^{-19}$ | <b>0.87</b> | 186,647 | $10^{-25}$ | <b>0.93</b> | 279,140 | $10^{-18}$ | <b>0.86</b> | 592,815 | 0.09       | –           |
| 2-Mul   | 19,853          | 0.36       | –           | 16,680  | 0.01       | 0.60        | 13,312  | $10^{-7}$  | 0.71        | 19,995  | 0.29       | –           |
| 3-Mul   | 955,418         | $10^{-3}$  | 0.63        | 678,403 | $10^{-11}$ | <b>0.77</b> | 591,748 | $10^{-22}$ | <b>0.89</b> | 975,558 | $10^{-4}$  | 0.65        |
| DeMux   | 19,633          | $10^{-5}$  | 0.68        | 16,678  | $10^{-12}$ | <b>0.79</b> | 29,700  | 0.59       | –           | 19,098  | $10^{-5}$  | 0.67        |
| Comp    | 542,290         | $10^{-3}$  | 0.63        | 453,730 | 0.44       | –           | 298,758 | $10^{-4}$  | 0.66        | 576,263 | $10^{-4}$  | 0.64        |
| 3-EP    | 6283            | 0.05       | –           | 5248    | $10^{-3}$  | 0.61        | 5990    | $10^{-3}$  | 0.61        | 5860    | 0.08       | –           |
| 4-EP    | 23,828          | 0.06       | –           | 20,278  | $10^{-5}$  | 0.66        | 18,745  | $10^{-6}$  | 0.69        | 20,295  | $10^{-3}$  | 0.62        |
| 5-EP    | 57,333          | 0.01       | 0.60        | 58,408  | $10^{-3}$  | 0.62        | 43,313  | $10^{-10}$ | <b>0.76</b> | 60,087  | 0.01       | 0.60        |
| 6-EP    | 129,910         | $10^{-5}$  | 0.67        | 134,770 | 0.03       | 0.58        | 104,392 | $10^{-9}$  | <b>0.74</b> | 113,037 | $10^{-6}$  | 0.68        |
| 7-EP    | 232,735         | $10^{-9}$  | <b>0.75</b> | 330,572 | 0.05       | 0.58        | 221,790 | $10^{-12}$ | <b>0.78</b> | 219,237 | $10^{-12}$ | <b>0.78</b> |

The  $p$  value is from the two-tailed Mann–Whitney  $U$  test. Where  $p < 0.05$ , the effect size from the Vargha–Delaney  $A$  test is shown; large effect sizes ( $A > 0.71$ ) are shown in bold

number of evaluations) required. Additionally, they offer strong evidence for the claim that there are some neutral rule-sets which may generally improve performance for a wide range of problems, particularly evidenced by the DMN and ID rule-sets.

We identify DMN and ID as the best performing rule-sets; each of these yield significant improvements in performance across all but one problems (the exceptions being Comp and DeMux, respectively), and on those single problems that they fail to improve upon, their inclusion does not lead to significant detriment in performance. For this reason, these rule-sets are the subject of further analysis in Sect. 8.

## 8 Analysis

### 8.1 Neutral drift or neutral growth?

Analysis of the runtime of EGGP augmented with the DMN and ID neutral rule-sets reveals their bias towards searching the space of larger solutions. When we refer to larger solutions, given that EGGP uses fixed-size representations, we refer to the proportion of the individual graph which is active, defined by the number of nodes to which there is a path from an output node. We demonstrate this with the results given in Table 5. Here, we measure the average (mean) size of the single surviving member throughout evolutionary runs on the 3-Add and Comp problems and give the median and interquartile range of

these average sizes over 100 runs. The size of an individual is the number of active function nodes (those which are reachable from output nodes) contained within it. We give these values for DMN, ID and EGGP alone. We use a two-tailed Mann–Whitney  $U$  test to measure for statistical differences between these observations. On both problems, DMN has a higher median average size (MAS) than both ID and EGGP alone ( $p < 0.05$ ) and ID also has a higher MAS than EGGP alone ( $p < 0.05$ ).

This observation challenges existing ideas that increasing the proportion of inactive code aids evolution (Miller and Smith 2006). We are able to achieve improvements in performance while effectively reducing the proportion of inactive code. It may be the case that high proportions of inactive code are helpful only when other forms of neutral drift are not available.

The result that DMN and ID increase the active size of individuals initially appears to challenge our hypothesis that it is semantic neutral drift that aids evolution. An alternative explanation could be that it is ‘neutral growth’, where our neutral rule-sets increase the size of individuals, that biases search towards larger solutions, which then happen to be better candidates for the problems we study. However, the CC neutral rule-set exclusively features neutral growth and neutral shrinkage, exploiting no domain knowledge beyond the notion that identical nodes in identical circumstances perform the same functionality, and featuring no meaningful semantic rewriting. We therefore compare how CC and DMN perform with different numbers of nodes available, to determine whether

**Table 5** Observed average solution size of the surviving population for the DMN rule-set, ID rule-set and EGGP without a neutral rule-set

| Problem | DMN  |      | ID   |     | EGGP |     | $p$           |                 |                |
|---------|------|------|------|-----|------|-----|---------------|-----------------|----------------|
|         | MAS  | IQR  | MAS  | IQR | MAS  | IQR | DMN versus ID | DMN versus EGGP | ID versus EGGP |
| 3-Add   | 96.9 | 1.3  | 92.3 | 1.2 | 50.8 | 2.6 | $10^{-33}$    | $10^{-34}$      | $10^{-34}$     |
| Comp    | 99.3 | 95.6 | 92.3 | 0.5 | 67.0 | 2.3 | $10^{-34}$    | $10^{-34}$      | $10^{-34}$     |

Results are for the 3-Bit Adder (3-Add) and  $4 \times 1$ -Bit Comparator (Comp) problems. For each result, the Median Average Size (MAS) and Interquartile Range (IQR) are given. The  $p$  value is from the two-tailed Mann–Whitney  $U$  test

larger solutions are indeed better candidates for the studied problems.

We run DMN, CC and standard EGGP on the 2-Add, 3-Add and Comp problems, with fixed representation sizes of 50, 100 and 150 nodes. If it is the case that larger solutions are better candidates, and that our neutral rule-sets bias towards neutral growth, then we would expect to see degradation of performance (more evaluations needed) with a size of 50, and improvements (fewer evaluations needed) with a size of 150, over a baseline size of 100.

The results of these runs are shown in Fig. 8. For 2-Add and 3-Add with the DMN neutral rule-set, performance actually degrades when increasing the fixed size from 100 to 150, while remaining relatively similar when decreasing the size to 50. For EGGP alone and for the CC neutral rule-set, performance remains relatively similar when increasing the fixed size from 100 to 150, but degrades when decreasing the size to 50. These observations imply that the DMN rule-set is not simply growing solutions to a more beneficial search space, since it performs better when limited to a smaller space. Therefore, on these problems, there is some other property of the DMN rule-set that is benefiting performance.

For the Comp problem, trends remain similar for EGGP alone and the CC neutral rule-set. However, the performance of the DMN rule-set degrades when the fixed size is decreased from 100 to 50. This suggests that the Comp problem is in some way different from the other problems. Further, when DMN is run on the Comp problem, the average proportion of active code is nearly 100%. This may offer an explanation to why the DMN rule-set struggles to outperform standard EGGP on the Comp problem, which has more than twice as many outputs (18) as the next nearest problem (8, DeMux). DMN's bias towards growth paired with the high number of outputs may give some of the problem's many outputs little room to change and configure to a correct solution.

## 8.2 DMN and ID in combination

We investigate the effect of using DMN and ID, our two best performing neutral rule-sets, in combination. This

combined set, which we refer to as DMID, consists of the following logical equivalence laws:

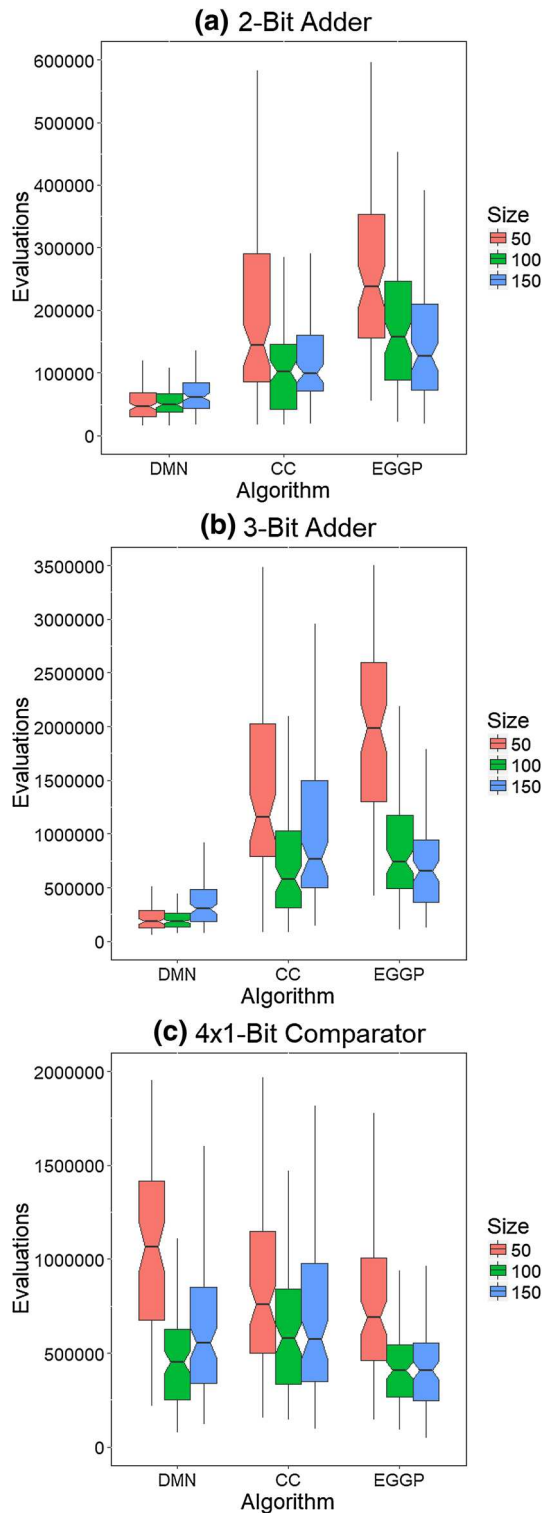
DeMorgan $_{F1}$ , DeMorgan $_{F2}$ ,  
 DeMorgan $_{R1}$ , DeMorgan $_{R2}$ ,  
 ID-AND $_F$ , ID-AND $_R$ ,  
 ID-OR $_F$ , ID-OR $_R$ ,  
 ID-NOT $_F$  and ID-NOT $_R$ .

We use this set under the same experimental conditions described in Sect. 6 to produce the results given in Table 6. In Table 6 we provide  $p$  and  $A$  values in comparison to the DMN and ID results in Table 4 and the EGGP results in Table 3.

The DMID rule-set significantly outperforms DMN on 7 of the 12 problems, and shows no significant difference for the other 5 problems. DMID significantly outperforms ID on 5 problems (notably the  $n$ -Bit Adder problems), shows no significant difference on 3 problems, and is significantly outperformed by ID on 4 problems (notably the 3-Mul, Comp and 7-EP). DMID significantly outperforms EGGP without neutral rule-sets on all but 1 problem, with the exception being the Comp problem that DMN also fails to find significant benefits on. These results position DMID and ID on a Pareto front of studied problems, with DMID effectively dominating DMN but neither DMID nor ID universally outperforming each other.

## 8.3 {AND, OR, NOT}: A harder function set?

In Table 3 we show that solving problems with the function set {AND, OR, NOT} is significantly more difficult than when using the function set {AND, OR, NAND, NOR}. We justify using the former function set over the latter in our experiments as it lends itself to known logical equivalence laws despite costing performance. When we introduce these logical equivalence laws to the evolutionary process with the {AND, OR, NOT} function set, this 'cost' no longer universally holds. We identify 3-Add, 3-Mul, Comp and 7-EP as the 4 hardest problems, based on the median number of evaluations required to solve them, Table 3. EGGP with the {AND, OR, NOT} function set and augmented with the



**Fig. 8** Results of running DMN, CC and EGGP on **a** 2-Add, **b** 3-Add and **c** Comp problems. The y axis gives the median evaluations required to solve each problem across 100 runs. The x axis groups setups by algorithm and then lists the observed median evaluations when running that algorithm with 50, 100 or 150 nodes as the fixed representation size

DMID neutral rule-set significantly ( $p < 0.05$ ) outperforms EGGP with the {AND, OR, NAND, NOR} function set on two of the problems.

These two are the 3-Add ( $p = 10^{-10}$ ,  $A = 0.76$ ) and 3-Mul problems ( $p = 10^{-5}$ ,  $A = 0.68$ ). In contrast, the reverse holds for Comp ( $p = 10^{-18}$ ,  $A = 0.85$ ) and 7-EP ( $p = 10^{-14}$ ,  $A = 0.80$ ). Note that for 3 of these circumstances (excluding 3-Mul), the significant difference occurs with large effect size ( $A > 0.71$ ).

Figure 9 shows the number of evaluations across 100 runs for the 3-Mul and Comp problems, for (A) EGGP with the {AND, OR, NOT} function set and augmented with the DMID neutral rule-set and (B) EGGP with the {AND, OR, NAND, NOR} function set. Here the difference in medians and interquartile ranges for these two evolutionary algorithms can be clearly seen; with EGGP with the DMID neutral rule-set requiring a median evaluations outside of the interquartile range of EGGP with the {AND, OR, NAND, NOR} function set for the 3-Mul problem. In stark contrast, the third quartile of evaluations required for the Comp problem lies below the first quartile of EGGP with the DMID neutral rule-set.

This offers an interesting secondary result: there are circumstances and problems where it may be beneficial to choose representations that on their own would yield detrimental results, if that decision then facilitates the inclusion of semantic neutral drift, which may in combination provide enhanced performance over the original representation.

### 9 Conclusions and future work

We have investigated the augmentation of a genetic programming system for learning digital circuits with semantic neutral drift. From our experimental results, we can draw a number of conclusions both for our own specific setting and for the broader evolutionary community.

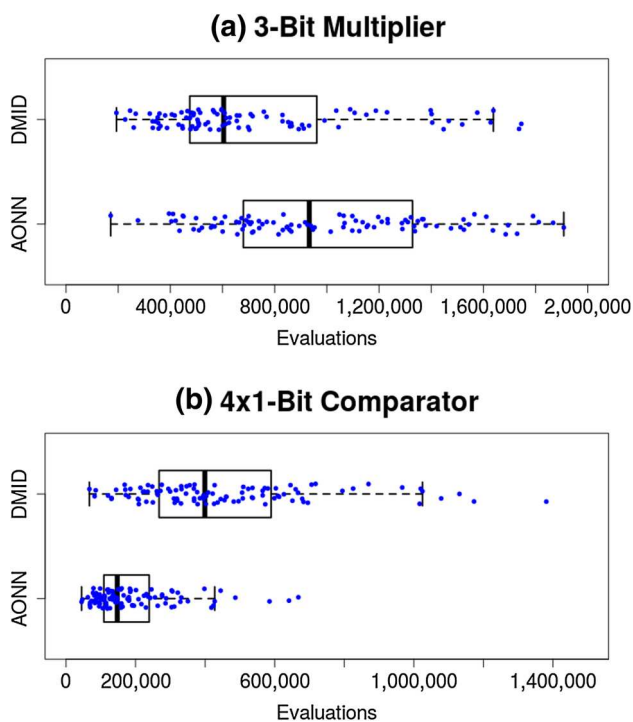
Firstly, we offer further evidence that there are circumstances where neutral drift aids evolution, building upon existing works that offer evidence in this direction. Additionally, the precise nature of our neutral drift by design offers evidence that neutral drift on the active component of individuals, rather than the intronic components, can aid evolution. For every benchmark problem studied, at least one neutral rule-set was able to yield significant improvements in performance.

Secondly, we have shown that by using graphs as a representation and graph programming as a medium for mutation, it is possible to directly inject domain knowledge into an evolutionary system to improve performance. The

**Table 6** Results from digital circuit benchmarks for the DMID neutral rule-set

| Problem | DMID    |         | versus DMN |      | versus ID  |             | versus EGGP |             |
|---------|---------|---------|------------|------|------------|-------------|-------------|-------------|
|         | ME      | IQR     | $p$        | $A$  | $p$        | $A$         | $p$         | $A$         |
| 1-Add   | 7415    | 5756    | $10^{-4}$  | 0.64 | 0.02       | 0.60        | $10^{-12}$  | <b>0.78</b> |
| 2-Add   | 43,633  | 29,065  | 0.13       | –    | $10^{-8}$  | <b>0.73</b> | $10^{-23}$  | <b>0.91</b> |
| 3-Add   | 162,568 | 112,074 | 0.02       | 0.60 | $10^{-11}$ | <b>0.77</b> | $10^{-28}$  | <b>0.95</b> |
| 2-Mul   | 12,020  | 8761    | $10^{-3}$  | 0.63 | 0.30       | –           | $10^{-8}$   | <b>0.73</b> |
| 3-Mul   | 604,480 | 471,956 | 0.51       | –    | 0.04       | 0.59        | $10^{-13}$  | <b>0.80</b> |
| DeMux   | 20,938  | 11,040  | $10^{-3}$  | 0.63 | $10^{-6}$  | 0.69        | $10^{-5}$   | 0.68        |
| Comp    | 399,140 | 315,459 | 0.45       | –    | $10^{-4}$  | 0.66        | 0.95        | –           |
| 3-EP    | 3930    | 3105    | $10^{-3}$  | 0.60 | $10^{-3}$  | 0.61        | $10^{-7}$   | 0.71        |
| 4-EP    | 16,778  | 10,730  | 0.02       | 0.59 | 0.13       | –           | $10^{-9}$   | <b>0.75</b> |
| 5-EP    | 52,868  | 31,445  | 0.29       | –    | $10^{-3}$  | 0.61        | $10^{-5}$   | 0.66        |
| 6-EP    | 121,978 | 90,429  | $10^{-3}$  | 0.61 | 0.11       | –           | $10^{-6}$   | 0.68        |
| 7-EP    | 326,040 | 224,121 | 0.95       | –    | $10^{-7}$  | 0.70        | 0.05        | 0.58        |

The  $p$  value is from the two-tailed Mann–Whitney  $U$  test. Where  $p < 0.05$ , the effect size from the Vargha–Delaney  $A$  test is shown; large effect sizes ( $A > 0.71$ ) are shown in bold. Statistics are given in comparison to the DMN and ID neutral rule-sets and EGGP



**Fig. 9** Box-plots showing observed evaluations required to solve **a** 3-bit multiplier and **b**  $4 \times 1$ -bit comparator problems using EGGP augmented with the DMID neutral rule-set (DMID) and EGGP with the {AND, OR, NAND, NOR} function set (AONN). Vertical jitter is included for visual clarity

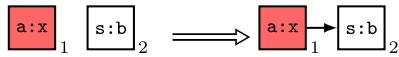
application of DeMorgan’s logical equivalence laws to graphs with sharing is non-trivial, but becomes immediately accessible in our graph evolution framework. Our ability to design complex domain-specific mutation operators supports the view that the choice of representation of individuals in an evolutionary algorithm matters. This injection of domain knowledge has been shown to offer benefits beyond simple ‘neutral growth’.

Thirdly, while the approach we have proposed here offers promising results, the specific design of neutral drift matters. There are neutral rule-sets that appear to dominate each other, as is found comparing the DMID rule-set to the DMN rule-set. There are also neutral rule-sets which outperform each other on different problems, as is demonstrated comparing the DMID rule-set to the ID rule-set. As we highlighted in comparing DMID to EGGP with what initially appeared to be a preferential function set, there are circumstances where a GP practitioner may want to deliberately degrade the representation in order to access beneficial neutral drift techniques. There are also other circumstances where the cost of incorporating these techniques may outweigh their immediate benefits.

There are a number of immediate extensions to our work that we believe should be investigated. Firstly, the use of the complete function set {AND, OR, NAND, NOR, NOT} alongside the DMID semantics preserving mutations and additional mutations for converting between AND and OR

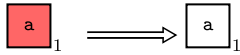
Main := ([{add\_node\_f<sub>x</sub> | f<sub>x</sub> ∈ F}]; [connect\_node]!; unmark\_node!); [connect\_output]!; remove\_counter

connect\_node(a,b:list; s:string; x:int)

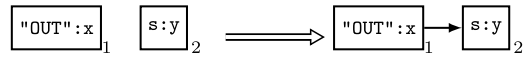


where s != "OUT" and outdeg(1) < x

unmark\_node(a:list)

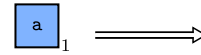


connect\_output(s:string; x,y:int)



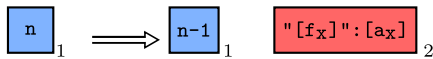
where s != "OUT" and outdeg(1) = 0

remove\_counter(a:list)



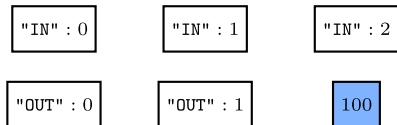
**Fig. 10** A P-GP 2 program InitCircuit for generating digital circuits. The program repeatedly probabilistically applies a add\_node\_f<sub>x</sub> rule (see Fig. 11 as long as possible, probabilistically connecting each newly added function node to the existing graph with the connect\_node rule until the node's function arity is satisfied. Once the add\_node rules are no longer applicable, the

add\_node\_f<sub>x</sub>(n:int)



where n > 0

**Fig. 11** A P-GP 2 rule for adding a node of some function f<sub>x</sub>. For the label of node 2 on the right-hand-side and a specific function f<sub>x</sub>, a unique string representation of f<sub>x</sub> replaces '[f<sub>x</sub>]' and the arity of f<sub>x</sub> replaces '[a<sub>x</sub>]'. The blue marked node counter is decreased, and the created function node is marked red so that its edges can be inserted. (Color figure online)



**Fig. 12** The initial graph to be used as input to the program in Fig. 10. Applying the program InitCircuit to this graph will generate acyclic graphs with 3 inputs, 2 outputs and 100 function nodes

gates and their negations via NOT should be investigated. It may be the case that this overall combination yields better results than either of the function sets and semantics preserving mutations we have covered in this work. Additionally, while semantics preserving mutations have generally improved performance with respect to the

connect\_output rule is applied as long as possible to connect the outputs to the rest of the graph. Finally remove\_counter cleans the graph up, removing the blue marked counter node. The generated graph must be acyclic, as edges are only created outgoing from nodes with no incoming edges. (Color figure online)

number of evaluations required to solve problems, it would be worthwhile to measure the clock-time cost of executing these transformations in every generation. Then it would be possible to study the trade-off between gained efficiency and additional overhead. Future work should also investigate the potential use of our proposed approach in CGP and tree-based GP as discussed in Sect. 5.1.

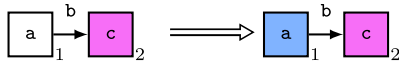
While we do not address theoretical aspects of SND here, it may be possible to prove convergence of evolutionary algorithms equipped with SND under certain properties, such as the completeness of the semantics preserving mutations used with respect to equivalence classes.

There are a number of application domains to investigate for future work: hard search problems where individual solutions may be represented by graphs and where there are known semantics-preserving laws. A primary candidate is the evolution of Bayesian Network topologies, a well-studied field (Larrañaga et al. 2013), as there are known equivalence classes for Bayesian Network topologies (Chickering 2002). A secondary candidate is learning quantum algorithms using the ZX-calculus, which represents quantum computations as graphs (Coecke and Duncan 2011), and is equipped with graphical equivalence laws that preserve semantics.

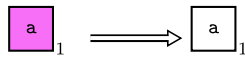


Main := [{mutate\_node\_f<sub>x</sub> | f<sub>x</sub> ∈ F}]; mark\_output!; [add\_edge,delete\_edge]!; unmark\_node!

mark\_output(a,b,c:list)

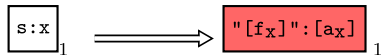


unmark\_node(a:list)



**Fig. 13** A P-GP 2 program MutateNode for mutating function nodes of digital circuits. The program probabilistically applies a mutate\_node\_f<sub>x</sub> rule (see Fig. 14 to mutate a node’s function and mark that node red. In a similar manner to the edge mutation program in Fig. 3, all nodes with a directed path to the mutating node are marked blue by mark\_output applied as long as possible. Then

mutate\_node\_f<sub>x</sub>(s:string; x:int)



where s != "IN" and s != "OUT" and s != "[f<sub>x</sub>]"

**Fig. 14** A generic P-GP 2 rule for mutating a function node to some function f<sub>x</sub>. For the label of node 1 on the right-hand-side and a specific function f<sub>x</sub>, a unique string representation of f<sub>x</sub> replaces ‘[f<sub>x</sub>]’ and the arity of f<sub>x</sub> replaces ‘[a<sub>x</sub>]’

**Acknowledgements** Timothy Atkinson is supported by a Doctoral Training Grant from the Engineering and Physical Sciences Research Council (EPSRC) Grant no. (1789003) in the UK.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

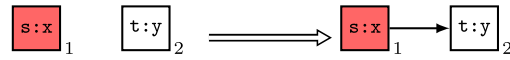
## Appendix: EGGP programs

### InitCircuit

The program InitCircuit in Fig. 10 generates EGGP individuals for the digital circuit problems described in this work. This program is defined abstractly, for some function set *F*. The actual form of the first rule-set call is instantiated for a specific function set *F* where each function f<sub>x</sub> has a corresponding version of the rule add\_node\_f<sub>x</sub> shown in Fig. 11.

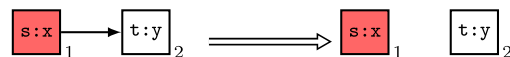
This program expects the a problem-specific variant of the graph given in Fig. 12, where there are *i* input nodes

add\_edge(s,t:string; x,y:int)



where t != "OUT" and outdeg(1) < x

delete\_edge(s,t:string; x,y:int)



where outdeg(1) > x

add\_edge and delete\_edge can be applied as long as possible to ensure that the node’s outgoing edge’s respect its new function’s arity. Additionally, the fact that all nodes which would introduce a cyclic if tareded are now marked blue ensures that applying add\_edge cannot introduce a cycle. Finally unmark\_node is used to return the graph to an unmarked state. (Color figure online)

and *o* output nodes and the blue node is labelled with *n* where *n* is an integer representing the number of nodes generated individuals should contain. The specific graph in Fig. 12 will generate circuits with 3 input nodes, 2 output nodes and 100 function nodes.

### MutateNode

The program MutateNode in Fig. 13 mutates EGGP individuals’ function nodes for the digital circuit problems described in this work. This program is defined abstractly, for some function set *F*. The actual form of the first rule-set call is instantiated for a specific function set *F* where each function f<sub>x</sub> has a corresponding version of the rule mutate\_node\_f<sub>x</sub> shown in Fig. 14.

## References

Atkinson T, Plump D, Stepney S (2018a) Evolving graphs by graph programming. In: Proceedings of 21st European conference on genetic programming (EuroGP 2018), LNCS, vol. 10781. Springer, pp 35–51. [https://doi.org/10.1007/978-3-319-77553-1\\_3](https://doi.org/10.1007/978-3-319-77553-1_3)

Atkinson T, Plump D, Stepney S (2018b) Probabilistic graph programs for randomised and evolutionary algorithms. In: Proceedings of 10th international conference on graph transformation, LNCS, vol. 10887. Springer, pp 63–78. [https://doi.org/10.1007/978-3-319-92991-0\\_5](https://doi.org/10.1007/978-3-319-92991-0_5)

Banzhaf W (1994) Genotype–phenotype-mapping and neutral variation—a case study in genetic programming. In: Proceedings of 3rd international conference on parallel problem solving from nature, LNCS, vol. 866. Springer, pp 322–332. [https://doi.org/10.1007/3-540-58484-6\\_276](https://doi.org/10.1007/3-540-58484-6_276)

Barnett L (1998) Ruggedness and neutrality; the NKp family of fitness landscapes. In: Proceedings of 6th international

- conference on artificial life. MIT Press, pp 18–27. <http://dl.acm.org/citation.cfm?id=286139.286143>
- Chickering DM (2002) Learning equivalence classes of Bayesian-network structures. *J Mach Learn Res* 2:445–498
- Coecke B, Duncan R (2011) Interacting quantum observables: categorical algebra and diagrammatics. *New J Phys* 13(4):043016. <https://doi.org/10.1088/1367-2630/13/4/043016>
- Collins M (2006) Finding needles in haystacks is harder with neutrality. *Genet Program Evolvable Mach* 7(2):131–144. <https://doi.org/10.1007/s10710-006-9001-y>
- Downing RM (2005) Evolving binary decision diagrams using implicit neutrality. In: Proceedings of 2005 IEEE congress on evolutionary computation, vol. 3. IEEE, pp 2107–2113. <https://doi.org/10.1109/CEC.2005.1554955>
- Downing RM (2006) Neutrality and gradualism: encouraging exploration and exploitation simultaneously with binary decision diagrams. In: Proceedings of 2006 IEEE congress on evolutionary computation. IEEE, pp 615–622. <https://doi.org/10.1109/CEC.2006.1688367>
- Galván-López E, Poli R, Kattan A, O'Neill M, Brabazon A (2011) Neutrality in evolutionary algorithms: What do we know? *Evol Syst* 2(3):145–163. <https://doi.org/10.1007/s12530-011-9030-5>
- Habel A, Kreowski HJ, Plump D (1988) Jungle evaluation. In: Recent trends in data type specification (WADT'87), selected papers, LNCS, vol. 332. Springer, pp 92–112. [https://doi.org/10.1007/3-540-50325-0\\_5](https://doi.org/10.1007/3-540-50325-0_5)
- Hahn MW (2007) Toward a selection theory of molecular evolution. *Evolution* 62(2):255–265. <https://doi.org/10.1111/j.1558-5646.2007.00308.x>
- Harvey I, Thompson A (1997) Through the labyrinth evolution finds a way: a silicon ridge. In: Proceedings of 1st international conference on evolvable systems, LNCS, vol. 1259. Springer, pp 406–422. [https://doi.org/10.1007/3-540-63173-9\\_62](https://doi.org/10.1007/3-540-63173-9_62)
- Hu T, Banzhaf W (2009) Neutrality and variability: two sides of evolvability in linear genetic programming. In: Proceedings of the 11th annual conference on genetic and evolutionary computation, GECCO '09. ACM, pp 963–970. <https://doi.org/10.1145/1569901.1570033>
- Hu T, Banzhaf W (2018) Neutrality, robustness, and evolvability in genetic programming. Springer, pp 101–117. [https://doi.org/10.1007/978-3-319-97088-2\\_7](https://doi.org/10.1007/978-3-319-97088-2_7)
- Husa J, Kalkreuth R (2018) A comparative study on crossover in Cartesian genetic programming. In: Proceedings of European conference on genetic programming, EuroGP 2018, LNCS, vol. 10781. Springer, pp 203–219
- Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge
- Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge
- Larrañaga P, Karshenas H, Bielza C, Santana R (2013) A review on evolutionary algorithms in Bayesian network learning and inference tasks. *Inf Sci* 233:109–125. <https://doi.org/10.1016/j.ins.2012.12.051>
- López EG, Rodríguez-Vázquez K (2007) Multiple interactive outputs in a single tree: an empirical investigation. In: Proceedings of 10th European conference on genetic programming (EuroGP 2007), LNCS, vol. 4445. Springer, pp 341–350. [https://doi.org/10.1007/978-3-540-71605-1\\_32](https://doi.org/10.1007/978-3-540-71605-1_32)
- Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 18(1):50–60
- Miller JF (ed) (2011) Cartesian genetic programming. Springer, Berlin. <https://doi.org/10.1007/978-3-642-17310-3>
- Miller JF, Smith SL (2006) Redundancy and computational efficiency in cartesian genetic programming. *IEEE Trans Evol Comput* 10(2):167–174. <https://doi.org/10.1109/TEVC.2006.871253>
- Miller JF, Thomson P (2000) Cartesian genetic programming. In: Proceedings of 3rd European conference on genetic programming (EuroGP 2000), LNCS, vol. 1802. Springer, pp 121–132. [https://doi.org/10.1007/978-3-540-46239-2\\_9](https://doi.org/10.1007/978-3-540-46239-2_9)
- Moraglio A, Krawiec K, Johnson CG (2012) Geometric semantic genetic programming. In: Parallel problem solving from nature—PPSN XII, LNCS, vol. 7491. Springer, pp 21–31. [https://doi.org/10.1007/978-3-642-32937-1\\_3](https://doi.org/10.1007/978-3-642-32937-1_3)
- Nguyen QU, Nguyen XH, O'Neill M (2009) Semantic aware crossover for genetic programming: the case for real-valued function regression. In: Proceedings of European conference on genetic programming, EuroGP 2009, LNCS, vol. 5481. Springer, pp 292–302. [https://doi.org/10.1007/978-3-642-01181-8\\_25](https://doi.org/10.1007/978-3-642-01181-8_25)
- Pham TA, Nguyen QU, Nguyen XH, O'Neill M (2013) Examining the diversity property of semantic similarity based crossover. In: Proceedings of European conference on genetic programming, EuroGP 2013, LNCS, vol. 7831. Springer, pp 265–276
- Plump D (2012) The design of GP 2. In: Proceedings of workshop on reduction strategies in rewriting and programming (WRS 2011), EPTCS, vol. 82, pp 1–16. <https://doi.org/10.4204/EPTCS.82.1>
- Plump D (2017) From imperative to rule-based graph programs. *J Log Algebraic Methods Program* 88:154–173. <https://doi.org/10.1016/j.jlamp.2016.12.001>
- Poli R (1997) Evolution of graph-like programs with parallel distributed genetic programming. In: Proceedings of international conference on genetic algorithms. Morgan Kaufmann, pp 346–353
- Poli R (1999) Parallel distributed genetic programming. In: New ideas in optimization. McGraw-Hill, pp 346–353
- Turner AJ, Miller JF (2015a) Introducing a cross platform open source cartesian genetic programming library. *Genet Program Evolvable Mach* 16(1):83–91. <https://doi.org/10.1007/s10710-014-9233-1>
- Turner AJ, Miller JF (2015b) Neutral genetic drift: an investigation using cartesian genetic programming. *Genet Program Evolvable Mach* 16(4):531–558. <https://doi.org/10.1007/s10710-015-9244-6>
- Vanneschi L, Pirola Y, Mauri G, Tomassini M, Collard P, Verel S (2012) A study of the neutrality of Boolean function landscapes in genetic programming. *Theor Comput Sci* 425:34–57. <https://doi.org/10.1016/j.tcs.2011.03.011>
- Vanneschi L, Castelli M, Silva S (2014) A survey of semantic methods in genetic programming. *Genet Program Evolvable Mach* 15(2):195–214. <https://doi.org/10.1007/s10710-013-9210-0>
- Vargha A, Delaney HD (2000) A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *J Educ Behav Stat* 25(2):101–132
- Vassilev VK, Miller JF (2000) The advantages of landscape neutrality in digital circuit evolution. In: Proceedings of 3rd international conference on evolvable systems, LNCS, vol. 1801. Springer, pp 252–263. [https://doi.org/10.1007/3-540-46406-9\\_25](https://doi.org/10.1007/3-540-46406-9_25)
- Walker JA, Miller JF (2008) The automatic acquisition, evolution and reuse of modules in cartesian genetic programming. *IEEE Trans Evol Comput* 12(4):397–417. <https://doi.org/10.1109/TEVC.2007.903549>
- Yu T, Miller J (2002) Finding needles in haystacks is not hard with neutrality. In: Proceedings of 5th European conference on genetic programming, LNCS, vol. 2278. Springer, pp 13–25. [https://doi.org/10.1007/3-540-45984-7\\_2](https://doi.org/10.1007/3-540-45984-7_2)
- Zhang J (2003) Evolution by gene duplication: an update. *Trends Ecol Evolut* 18(6):292–298. [https://doi.org/10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8)